Leveraging Previously Reviewed Documents with TAR Workflow
Managed Review – Case Study & Cost Analysis
Innovative Discovery, LLC

Context

The Client needed to review 150,112 documents for responsiveness and privilege with limited resources available. Taking into account a firm 10-day timeline, Innovative Discovery, LLC ("ID") created a workflow incorporating Technology Assisted Review ("TAR") to efficiently review the necessary documents. **Using the Analytic Tools listed below and an optimized work flow enabled the client to meet the production deadline while achieving significant time and cost savings – over $150,000 – when compared to a linear review.**

Challenges

Multiple factors, including a strict deadline, contributed to the complexity of this document review.

- A small portion of the review universe had been previously reviewed by Associates of the client firm, under a different protocol. Rather than re-reviewing these documents, ID leveraged this existing work product by selecting "Hot" documents and repurposing into training examples for the Seed Set during Categorization.

- Due to the time constraint, the client could not conduct the multiple training rounds necessary to reach peak accuracy. In light of this, ID conducted extensive verification sampling to mitigate the risk of inaccurate decisions made by the Categorization algorithm.

- The resources required to complete the review increased from the original estimate as new data was loaded into the workspace. ID quickly assessed the increase in total review hours and trained the appropriate number of additional legal reviewers to ensure the review was completed on schedule.

Analytic Tools

The following analytic tools[i] were used in this case study:

Categorization:
A search and organizational method whereby documents are categorized based on an example document or a set of example documents. This method requires user input in that the user must identify the exemplar document(s) before the categorization can take place.

Email Threading:
An organizational method whereby emails are grouped according to a single email conversation that starts with an original email (the beginning of the conversation), and is grouped with all of the subsequent replies and forwards pertaining to that original email.

Clustering:
An organizational method whereby documents are segregated into mutually exclusive groups, or "clusters," of conceptually similar[ii] documents based on similar text patterns within the documents. Clusters can be created without requiring any user input and can be implemented as soon as the documents are indexed.

Cost Savings Linear Review vs. TAR Review

Review Category	Documents per Hour	Docs. Reviewed	Hours to Complete Review	Cost to Complete Review
SME Review (Law Firm Associates)	60	1,374	22.9	-
Responsiveness Review (Including Second Level Review & Verification Sampling)	65	107,053	1,701.9	-
Privilege Review (Including Redactions & Privilege Logging)	35	24,222	692.1	-
Cost of Analytics (per GB)	N/A	N/A	269.9 GB	-
Project Management & TAR Consultation	N/A	N/A	359.1	-
TOTAL - TECHNOLOGY ASSISTED REVIEW			**2,775.9**	**$ 204,003.69**
Linear - Responsiveness Review (Including Second Level Review)	45	150,112	3,335.8	-
Linear - Privilege Review (Including Redactions & Privilege Logging)	15	30,182	2,012.1	-
Project Management	N/A	N/A	802.2	-
TOTAL - LINEAR REVIEW			N/A	**$ 364,997.97**
**SAVINGS USING TECHNOLOGY ASSISTED REVIEW				**$ 160,994.28**

Summary of TAR Workflow

At the beginning of the TAR review, the total number of documents in the Analytics Review Universe was **150,112**.

The Subject Matter Experts ("SMEs") reviewed a total of **1,374** documents that were used in the Seed Set during Categorization. After the target Recall was achieved, the total number of documents categorized as Responsive (including Family Members) was **89,982**. Due to the time constraint and reduced number of training rounds during Categorization, ID conducted extensive verification sampling to ensure algorithm accuracy. **10,127** documents went through verification sampling.

To identify potentially privileged documents, ID ran a privilege screen against the Potential Production Universe. The result was **24,222** documents.

Once Privilege Term Screening & Culling were completed, the resulting TAR review universe was:

107,053 **TAR – Total Documents requiring Human Responsiveness Review (Including Second Level Review)**
           89,982 (Documents categorized as Responsive)
           10,127 (Verification Sampling)
           6,944 (Second Level Review)

24,222 **TAR – Total Documents requiring Human Privilege Review (Including Redactions & Privilege Logging)**

50,003 **TAR – Total Documents NOT requiring Human Review**

---

[i] The analytic tools detailed in this document are available via kCura's web-based application, Relativity.

[ii] Relativity Analytics identifies documents with similar conceptual content, rather than matching specific search terms or character strings like traditional search engines. The Analytics engine uses a mathematically-based technology called Latent Semantic Indexing ("LSI") to identify relationship patterns between terms and concepts without using pre-established word lists. Instead, LSI leverages mathematics to discover term correlations and conceptuality within a document set. Thus, Relativity Analytics derives the potential meaning of a group of text by understanding the semantic meaning and the context of the terms in relation to the other terms present in the data set.